# Prédiction of diabete

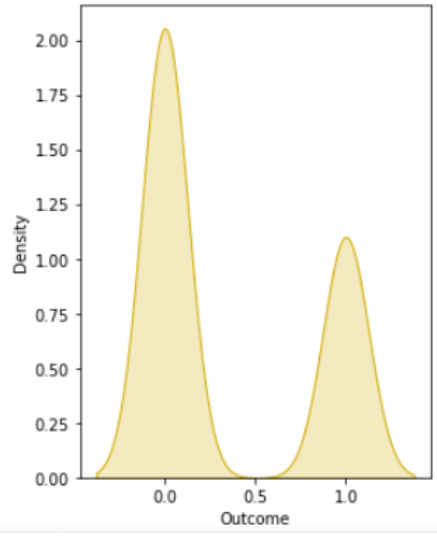
Diabetes is a chronic medical condition characterized by high levels of blood sugar (glucose). It occurs when the body either does not produce enough insulin or is unable to effectively use the insulin it produces. Insulin is a hormone produced by the pancreas that helps regulate blood sugar levels and allows cells to utilize glucose for energy.

## Summary

As a data analyst intern at MeriSkill, my project on diabetes prediction. The dataset is provided by MeriSkill. After training and testing the model, we retain the random forest as the best since it gives us an accuracy of 78% and an exact prediction to that which is real.

## Data mining:

The dataset has nine (09) columns including height (08) features and one (01) target. The target is the Outcome column. The target is composed of values 1 and 0, i.e. diabetic and non-diabetic. So, we made a graphical representation.




## Model used:

1. Logistics Regression

Accuracy = 0.7265625
Error = 0.2734375

2. KNeighbors Classifier

Accuracy = 0.7578125
Error =0.06993006993006994

3. Decision Tree Algorithm

Accuracy = 0.703125
Error =0.296875

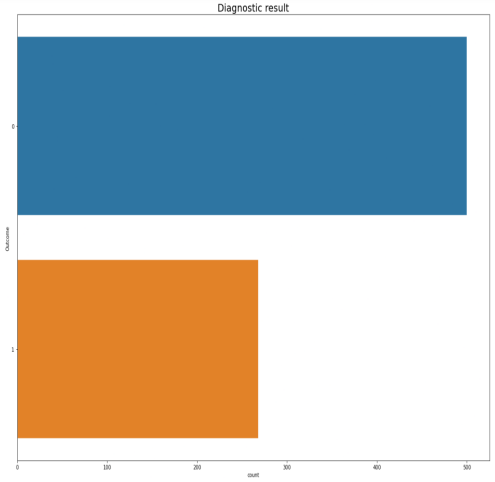
4. Random Forest Classification algorithm

Accuracy = 0.7890625
Error =0.2109375

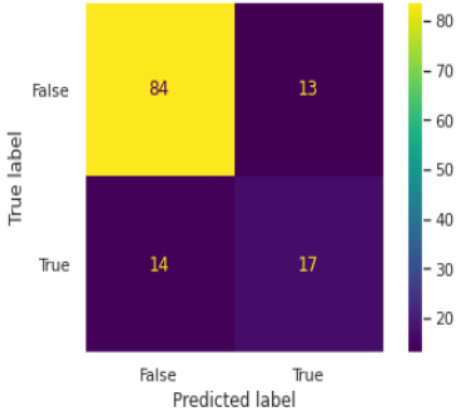
5. XGBOOST

Accuracy = 0.7734375
Error =0.2265625

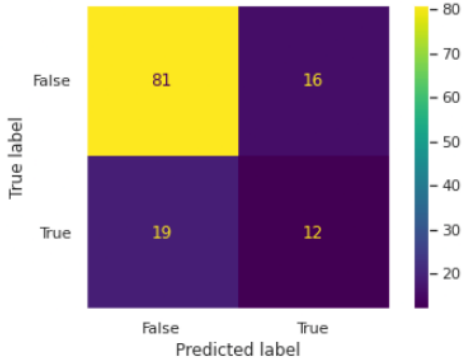
## RESULT




### Cas 2: Random Forest Classification




### Case 1 : Logistic Régression




## Conclusion :

This project is part of my internship as a Data Analyst at MeriSkill. The dataset used is that provided by the internship institution. We used several prediction algorithms. The model that gave better accuracy is the RandomForest Classifier, with an accuracy of __78%.__

We deployed the solution on
If you have any comments or suggestions, please send them to us.

## Github Link :

CREATED WITH: EDIT.org